2009

Developing New Indices for the Identification of Poor Effort

Scott A. Magnuson
Cleveland State University

DEVELOPING NEW INDICES FOR THE IDENTIFICATION

OF POOR EFFORT

SCOTT A. MAGNUSON

Bachelor of Arts in Psychology

University of Dayton

May, 2006

submitted in partial fulfillment of requirements for the degree

MASTER OF ARTS IN PSYCHOLOGY

at the

CLEVELAND STATE UNIVERSITY

May, 2009

This thesis has been approved

for the Department of PSYCHOLOGY

and the College of Graduate Studies by

Thesis Chairperson, Dr. Amir Poreh, Ph.D.

Department & Date

Committee Member, Dr. Conor T. McLennan, Ph.D.

Department & Date

Committee Member, Dr. Boaz Kahana, Ph.D.

Department & Date

DEVELOPING NEW INDICES FOR THE

IDENTIFICATION OF POOR EFFORT

SCOTT A. MAGNUSON

ABSTRACT

The purpose of this study is to gather data in order to develop new indices to aid clinicians in more accurately distinguishing between patients feigning posttraumatic stress disorder (PTSD) and those with genuine PTSD. Participants were administered a battery of tests that included the Digit Span, Digit Symbol, Trail Making Test (TMT), and Rey Auditory Verbal Learning Test (RAVLT) twice, once while performing genuinely and again while simulating PTSD. The results of this study found that trails 21-25 of part A of the TMT are a good indicator of poor effort. This measurement was found to have high sensitivity (90.7%) and specificity (82.2%). Construct validity was established with correlational analysis. Trails 21-25 of part A of the TMT highly correlated with several well validated measurements of effort such as the Reliable Digit Span (-.492) and the RAVLT Forced Choice Thirty Minute Delay (-.563).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

There are occurrences of individuals intentionally deceiving neuropsychological tests in order to appear more impaired than they truly are. The intentional deception of these tests is done for many reasons: financial compensation, avoiding work, obtaining drugs, receiving less severe criminal sentences, or simply to attract attention or sympathy. This deception is referred to as Malingering. The Diagnostic and Statistical Manual classifies Malingering as a V-code: behavior that may be worthy of clinical attention but not a mental disorder in itself. The DSM-IV-TR also states that the essential features of Malingering are "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (American Psychiatric Association, 2000, p. 739). A malingerer can feign many different psychological disorders and the motive to do so is broad.

Calculating prevalence rates for Malingering is difficult for obvious reasons. A survey of forensic evaluators revealed that base rates of malingering were estimated to occur in 15.7% of forensic evaluations and 7.4% in nonforensic clients (Rogers, Sewell & Goldstein, 1994). Another study found that 20%-30% of personal injury claimants were feigning PTSD in an attempt to receive financial compensation (Lees-Haley, 1997). These rates have astounding implications on the social, legal, economic, and health care fields. Absenteeism decreases productivity of industry; frivolous claims reduce social security, disability, worker's compensation and insurance benefits; and unwarranted patients deplete the medical system of costly resources. A 1997 study by Resnick stated that 40% of the PTSD patients in their sample considered totally disabled and receiving benefits demonstrated no disability. According to the Washington Post, health care fraud costs taxpayers more than $60 billion each year (Johnson, 2007). As evidenced, Malingering affects society as a whole in many ways.

Some disorders are more easily malingered than others are. PTSD seems to be one of the most common disorders malingered. This is partially because criteria for the disorder are based on subjective experiences. Therefore, diagnosis of this disorder relies completely upon self-report from the patient. The DSM-IV-TR states six criteria, listed below, that are required in order to diagnose PTSD (American Psychiatric Association, 2000).

A. The person has been exposed to a traumatic event
B. The traumatic event is persistently reexperienced
C. Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness
D. Persistent symptoms of increased arousal (not present before the trauma)
E. Duration of the disturbance is more than one month
F. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning

Clients may completely fabricate their symptoms, exaggerate the extent of their injuries, or erroneously attribute preexisting problems to the trauma (Guriel & Fremouw, 2003). Although the DSM-IV-TR recommends that clinicians verify traumatic events with documentation such as hospital records, police reports, military records, etc., clinicians do not always do this. Because symptoms of PTSD are so easily feigned, the DSM-IV-TR suggests clinicians explore certain scenarios that might serve as indicators for people feigning PTSD. Such indicators include if the examination is in the medico-legal context, if there is distinct discrepancy between the client's claimed distress and the objective findings, if the client is not cooperative with clinician's instructions either during or after the examination, or if the client has been diagnosed with Antisocial Personality Disorder (American Psychiatric Association, 2000). In addition, Pankratz and Binder (1997) identify six behaviors that clinicians should be aware of, as they are highly suggestive that clients may be malingering. These six behaviors include (1) marked inconsistency between observed and reported symptoms; (2) marked inconsistency between diagnosis and neuropsychological findings; (3) resistance, avoidance, or bizarre responses on standardized tests; (4) failure on specific measures of faking; (5) functional findings on medical examination; and (6) late onset of cognitive complaints following accident.

It is sometimes not obvious why someone malingering PTSD would seek clinical treatment. In most cases, the individual is seeking compensation and is required to prove in a court of law that financial reparation is necessary. They are often recommended by a lawyer to seek clinical treatment in order to obtain documentation that could potentially confirm their claim (Taylor et al., 2007). In a 2003 article by Frueh et al., clinicians reported, “many patients acknowledge using treatment sessions as a means of documenting, through clinicians’ progress reports, their psychiatric difficulties” (p. 89). Another common motive for the PTSD malingerer to seek treatment is to appear irremediable, thereby increasing not only the likelihood of collecting damages, but also the amount of money awarded.

Three distinct patterns of malingering have been identified by Resnick. The first is the pure malingerer, which is one who completely fabricates symptomology. The second type is the partial malingerer, commonly characterized by over reporting current symptoms, or continuing to report latent symptoms. The final type of malingering is false imputation. This is when real symptoms are intentionally misattributed to a traumatic event (1997). For the purposes of this study, we will be concentrating specifically on the pure malingerer. It is important for clinicians to be aware of each subtype of malingerer in order to more readily identify symptom feigning. Currently, there are only two absolute methods for identifying malingering: if the individual confesses, or if the individual is observed performing “normal” behavior that was previously claimed incapable (Hall & Hall, 2006). If a clinician suspects a client is malingering, it is recommended that the clinician refer the client for neuropsychological testing.

Research on psychometric testing for the purpose of identifying malingerers has been performed, but is limited in scope. The focus of this paper will be on the Digit Symbol Test, Digit Span Test, Rey Auditory Verbal Learning Test, and the Trail Making Test as tools for identifying individuals malingering PTSD. However, the tests will not be administered in the traditional format. These tests will be administered utilizing specialized computer software developed by Quantified Process Scoring Systems. This software was chosen because it will assist the researchers in collecting additional data that would otherwise not be obtained if the tests were administered in the traditional method. The software was designed to utilize the principles outlined in the Quantified Process Approach (Poreh, 2000). One of these principles is particularly important to this research. The Decomposition Paradigm is an "approach that investigates the relationship between test items of a given measure according to underlying facets, resulting in the development of new subscores" (Poreh, 2000; Pg.5). Decomposing these tests will not only allow researchers to develop new indices for the purpose of identifying people feigning PTSD, it will also provide further evidence for the Quantified Process Approach.

The purpose of this study is to develop new indices of effort in order to detect people feigning PTSD. Developing these indices can assist clinicians both in the medical and forensic setting to distinguish people feigning the disorder from those who genuinely have the PTSD. In doing so, money and resources can be more appropriately allocated to those who are genuinely in need of help and support.

1.1 The Digit Span

The Digit Span is most commonly known as one of the 13 subtests in the Wechsler Adult Intelligence Scale (Kaufman, 1999). However, administration of the Digit Span can be traced back to research studies in 1887 by Galton and Jacobs. These studies claimed that the Digit Span Test measured 'prehension,' a term that Jacobs coined and defined "as the mind's power of taking on certain material" (p. 79). Jacobs theorized that people's 'span of prehension' was a determinant of their mental capacity. Alfred

Binet and Theodore Simon later used the Digit Span in 1905 when they published their "measuring scale of intelligence." Validity was demonstrated by the test's ability to distinguish between normal and cognitively impaired individuals (Binet & Simon, 1905). The Binet-Simon became widely used throughout Europe and was later popularized in America when psychologist Henry Goddard learned of its existence and had it translated into English (Zenderland, 1998).

The Digit Span is a two part test. The first section is Digits Forward, which is administered by presenting the test taker with auditory numbers. The test taker is to repeat back the numbers in the exact order with which they were presented. As the test progresses, additional numbers are presented during each trial. Similarly, the second section of the test, Digits Backwards, begins by presenting the test taker with auditory numbers. However, the test taker is to repeat the numbers back in the exact opposite order in which they were presented. Traditionally, this test is administered by an examiner speaking the digits at a constant rate and then recording the test taker's response. However, in this study the test was administered utilizing a specialized computer program. The software presents the digits through auditory speakers at a constant rate. Presenting the digits at a precisely, constant rate is very important because this reduces test administrator error and variability. The test taker will repeat back the numbers as the examiner types exactly what is repeated. By utilizing this computer program, researchers will be able to identify the different types of errors that test takers make. Examples of such errors are insertions, reversals, repetitions, and omissions. An insertion is when the test taker repeats back a number that was not presented; a reversal is when a test taker switches two numbers in the series of numbers that were presented; a repetition is when the test taker mistakenly repeats a number; and an omission is when a number is erroneously omitted from the sequence.

The Digit Span is often perceived by patients as measuring memory; however, it is a more effective measure of attention and concentration (Tombaugh, 1996). Studies have shown that Digit Span Forward performance ability is relatively preserved in patients with brain and mental disorders. Therefore, a patient's extreme inadequate performance is often indicative of exaggeration of symptoms. In general, the normal range for maximum digit span forward has been found to be 7 +/- 2 (Miller, 1956). Although some believe that this range is too broad, this is consistent with findings.

Many different indices derived from the Digit Span have been used in the detection of incomplete effort: the Digit Span Forward, Digit Span Backward (both described above), and the Reliable Digit Span. The Reliable Digit Span is calculated by summing the longest forward and backward digit strings in which both trials were correct. A Reliable Digit Span score of seven or less has been found to have a false-positive error rate of 10% or less in nonmalingering brain injured patients, criminal forensic subjects, healthy samples, and people in significant clinical pain. A Reliable Digit Span score of less than six is extremely rare, and thus is likely to be indicative of poor effort (Heinly, Greve, et al., 2005).

A 2004 study implementing an archival research design utilized cases from three different groups. Twenty nine people had traumatic brain injury, 36 people were classified as probable malingerers based on specific criterion, and 22 people were classified as nonlitigating mild traumatic brain injuries. Each group was administered the WAIS-III. The probable malingering group obtained mean scores significantly less than the other two groups on the Digits Forward, Digits Backward, and Reliable Digit Span. These extreme low scores are statistically very rare, and thus, cause suspicion of incomplete effort. Overall, these findings suggest that scores derived from the Digit Span are accurate measures of effort (Axelrod et al., 2004).

Digit Span performance has never been shown to be affected in patients who have PTSD. However, people who feign the disorder attempt to be perceived as extremely impaired. They do not understand the nature of PTSD, and therefore assume that a person with this disorder will be unable to obtain an average score on this test. Therefore, it is expected that participants simulating PTSD will grossly exaggerate, or overplay the symptoms associated with PTSD. In doing so, researchers hypothesized that PTSD simulators will perform significantly worse on all indices of the Digit Span. More specifically, it is expected that this group will obtain a score of seven or less on the Reliable Digit Span.

1.2 The Digit Symbol

The Digit Symbol-Coding, like the Digit Span, is one of the 13 subtests in the Wechsler Adult Intelligence Scale. Dr. Jastrow of the University of Wisconsin originally devised the Digit Symbol-Coding Test as a measure of association learning (Starch, 1911). The test was later revised by Woodworth and Wells (1911) and ultimately was adopted by Wechsler.

Administration of the Digit Symbol-Coding begins with instructions read verbatim by the examiner to the test-taker. Following three demonstrations, the examinee is instructed to copy four practice symbols on an answer sheet. If done correctly, administration commences. At the top of the page, a key matches numbers one through nine to a specific symbol (i.e. O, X, +, =). The test taker has 90 seconds to match 90 digits to their appropriate symbol. Traditionally, test takers are instructed to fill in answers on a response sheet while the examiner keeps time using a stopwatch. However, for this study, the examiner will operate a computer and press a button every time the test taker fills in the correct answer. By examining relationships between test items, researchers will be provided with additional data. This new data can then be explored in order to discover underlying facets that have previously gone unnoticed. These underlying facets will be used in order to develop new indices. This is an example of the Decomposition Paradigm in the Quantified Process Approach (Poreh, 2000).

In a 2006 study, the Digit Symbol-Coding Test was administered to 103 college students and data were collected every 30 seconds. Researchers hypothesized that participants would obtain average or above average scaled scores, as well as exhibit a learning curve over succeeding intervals of the test. The groups' average score fell at approximately the 75^{th} percentile (M = 11.88, SD = 2.48), However, the hypothesized learning curve only occurred in 2.9% of the participants (Ryan, Schnakenberg-Ott, & Brown, 2004; Ryan & Kreiner, 2006). By decomposing this test, researchers were able to conclude that their hypothesis was incorrect and that their participants did not display a learning curve over successive trials. Thus, employing the use of specially designed computer software will allow researchers to obtain additional information that would otherwise be unattainable.

Raw scores of the Digit Symbol-Coding test are converted into scaled scores by referencing Tables A.1 and A.2 in the WAIS-III Administration and Scoring Manual.

Scaled scores range from 1-19 (M=10, SD=3). Examination of scores indicates that performance on this test is highly influenced with education level, gender, ethnicity (Kauffman & Lichtenberger, 2002) and age (Ryan et al., 2000). The Digit Symbol-Coding scaled score has been found to be the most powerful of the 11 subtests of the WAIS-R for discriminating between neurologically impaired seniors and healthy individuals. A 2005 study compared scores on the WAIS-R of 54 healthy individuals to a demographically matched sample of neurologically impaired seniors. This study resulted in an 83.3% sensitivity rate and a 79.6% specificity rate (Ryan, Brown, and Paolo, 2005).

Researchers hypothesized that participants would complete significantly fewer symbols when simulating PTSD compared to their baseline performances. It was believed that the participants simulating the disorder would attempt to show that the disorder has markedly affected their ability to perform on this test. However, literature shows that this disorder has no significant affect on Digit Symbol performance.

1.3 The Trail Making Test (TMT)

The Trail Making Test (TMT) was initially developed and used as a subtest in the Army Individual Test Battery (U.S. War Department, Adutant General's Office, 1944), but has become more widely known by its inclusion in the Halstead-Reitan Neuropsychological Battery (Reitman & Wolfson, 1985). The TMT is one of the most widely used neuropsychological tests, used by 87% of neuropsychologists surveyed (Sellers & Nadler, 1992). This timed test measures sustained visual attention, visual scanning, sequencing and cognitive flexibility (Hebben & Milberg, 2002). Overall, this is a measure of different components of executive functioning.

There are two parts to the TMT. The first part, known as TMT-A, requires to test taker to connect numbered circles as quickly as possible, going from 1-25, without picking their pencil up from the paper. If any errors are made, the examiner is required to instruct the test taker to continue from the point of error. Once completed, the test taker is given the second section of the examination, the TMT-B. This portion is very similar to TMT-A, but instead of connecting numbers 1-25, the page contains circles that are numbered 1-13 and lettered alphabetically from A-L. The test taker is to connect the circles in an alternating fashion, starting with the first number ("1"), connecting it to the first letter ("A"), and then connecting that to the second number ("2"). The examinee is to continue in such a fashion until the conclusion of the test.

In this study, the TMT will be administered utilizing computer software. The examinee responds on a paper answer sheet. The examiner will press a button every time the test taker connects one of the dots. This will enable the researcher to obtain data that would otherwise not be obtainable. This computer software allows the researchers to decompose the test in order to obtain additional subscores. This allows the researcher to identify any areas of the test that might illicit a different response from a particular clinical group.

Empirically, the TMT has been found to be sensitive to a wide variety of neuropathological conditions. However, research indicates that Part B is more sensitive in assessing cerebral dysfunction (Spreen & Strauss, 1998). In addition, if the amount of time to complete part B is much greater than the time to complete part A, it is indicative that the test taker might have difficulties with complex conceptual processing (Lezak, 1995).

Additional indices that emanate from the TMT have been widely researched. One of the more widely used is the TMT ratio scores. It is hypothesized that ratio scores are more sensitive to cognitive impairment than the original TMT scores because it uses the examinees as their own control (Martin, Hoffman & Frank, 2003). The ratio score is calculated by dividing the amount of time to complete TMT-B by the amount of time required to complete TMT-A. Goebel (1983) examined the ratio scores of brain-injured

TABLE I: TMT Ratio Score results

Experimental Group	**TMT-A**	**TMT-B**	**Ratio**
Simulating Malingerers	38 s	80 s	2.1
Brain Damaged	102 s	318 s	3.1

(Goeble, 1983)

individuals, individuals instructed to malinger a brain injury, and a control group. Data from this study are listed in Table I. Results from the study suggest that malingerer's ratio scores were typically smaller than that of individuals with genuine brain disorders. This occurrence can be interpreted as the malingerers not realizing how the two sections differ, and do not understand that there are much more complicated mental processes that are required for TMT-B.

Researchers hypothesized that participants simulating PTSD would perform significantly slower, and therefore require more time to complete all of the trails than when they were instructed to perform genuinely. It was believed that participants simulating the disorder would be unaware that TMT performances are not significantly affected by PTSD. Being unaware of these findings, researchers believed that they would perform significantly worse under simulation conditions compared to their baseline performances.

1.4 The Rey Auditory Verbal Learning Test (RAVLT)

The original Auditory Verbal Learning Test was originally developed by Claparède (1919), but has since been modified by Rey into its more popular form today (Boake, 2000). The RAVLT measures verbal learning, memory, proactive and retroactive interference, retention, encoding, and retrieval (Hebben & Milberg, 2002). This test requires the test taker to recall 15 words over five presenting trials. A second distracter list of 15 different words is then presented and recalled by the test taker. Following this, the test taker is asked to recall the original 15 words. The next trial is then administered after a 30 minute distraction period in which the test taker is again asked to recall the original 15 words. Next, the test taker is presented with one word at a time and is to identify whether or not it was read to them during the original first five trials. The final stage of the RAVLT presents two words to the test taker, who is instructed to distinguish which of the two words were read during the first five trials of the test. For the purposes of this study, the RAVLT will be administered utilizing computer software. The software will present all directions and stimuli in a precise and standardized manner. It will also allow researcher to investigate new indices utilizing the principles outlined in Poreh's Quantified Process Approach.

Performance on the RAVLT is calculated by comparing scores to demographically matched norms across specific trials. Demographics such as age, sex, and education have been found to affect performance. A study conducted in 1992 on memory functioning found that total number of words recalled across all trials decreased significantly with age and was positively correlated to education in 161 healthy adults aged 62 to 100 (Petersen, Smith, Kokman, Ivnik, & Tangalos). In another study comparing aged matched dementia patients to a control group, females in both groups out performed males, indicating that gender is an important variable as well (Kraiuhin et al., 1986). Therefore, it is important for researchers to utilize demographically sensitive norms when referencing scores.

Research has found that non-credible performance can be identified by examinees scoring significantly below demographically matched peers. Many indices have been developed from the RAVLT, each contributing uniquely to the overall diagnostic process. The first index that can be interpreted, which emerged from the Satellite Testing Paradigm (Poreh, 2000) is trial 1. Believed to reveal simple attention span, this index is easily interpreted. It is often compared to Forward Digit Span raw scores in order to determine fluctuations in attention. Typically, the tests are within two points of each other. However, Lezak (1995) noted that when Forward Digit Span scores exceed RAVLT trial one scores by more than two points, it could be due to an individual feeling overwhelmed with 15 words. In contrast, if RAVLT trial one scores are much higher than Forward Digit Span scores, it could be due to fluctuations in motivation (See also Woodard, 2006). Thus, if a participant's Forward Digit Span and RAVLT trial one scores significantly vary, this may be an indicator of inconsistent or suspect motivation.

Another index that may prove useful in identifying suspect effort is the Overall Learning Index, which compares relative performance across trials. By comparing the highest and lowest trials, data are obtained that reflect the extent to which an individual can benefit from repeated exposure to the word list. (Woodard, 2006). Moritz et al. administered the test to healthy controls, patients diagnosed with schizophrenia, and patients diagnosed with depression. Results showed that there was no significant difference in performance across groups (2001). In another study by Shum et al., patients who had sustained a traumatic brain injury were compared with age-matched control participants. The ratio of the number of words recalled after the best and worst trial performance was not found to be significantly different across groups despite significant group differences in the total number of words recalled (2000). This demonstrates that despite the potential diagnoses, the majority of people will display a learning curve over successive trials.

Studies have shown that the addition of a second delay recall and recognition at 60 minutes, known as the AVLTX, may be particularly useful in identifying inconsistent effort. In a 2004 study, Barrash et al. constructed an exaggeration index (EI) based on the notion that malingerers would have difficulty keeping track of the words that were successfully recalled on the 30 minute delay recall portion of the test and 60 minute recall portion. It is expected that words that were recalled during the 30 minute delay, words that were unquestionably learned, would be recalled during the 60 minute delay. The EI reveals memory performance that is characteristically inconsistent of brain damaged individuals. It encompasses seven aspects of improbable performance: exceedingly poor learning, lack of primacy effect, worsening recall, worsening recognition, failure to recognize learned words, failure to recognize recalled words, and exceedingly poor recognition (See Table II). An EI score cutoff of three is recommended, and studies using simulating samples found that the AVLTX was modestly sensitive and relatively specific to malingerers, yielding a 72% sensitivity rate (Suhr et al., 2004). Thus, this is a very accurate measure of effort.

TABLE II: Operational Definitions and Scaling of Inconsistencies Comprising the Exaggeration Index

Inconsistency operational definition	Scaled inconsistency score			
	0	1	2	3
Exceedingly poor learning Immediate recall of 15 word list, summed across learning Trials 1–5	≥28	23–27	18–22	0–17
Lack of primacy effect Immediate recall of words 1–3, summed across learning Trials 1–5	≥4	1–3	0	–
Worsening recall Delayed recall at 30 min minus delayed recall at 60 min	0–2	3	4	≥5
Worsening recognition Delayed recognition at 30 min minus recognition at 60 min	0–2	–	3	≥4
Learned words not recognized This inconsistency is scored only if the false negative responses cannot be attributed to a "nay-saying" bias (i.e., score if pt made two or more false positive responses): (a) words recalled on 4+ learning trials, but "not recognized" at 30 min, plus (b) words recalled on 4+ learning trials, but "not recognized" at 60 min	0–2	3–4	5–8	≥9
Recalled words not recognized (a) words recalled at 30 min, but "not recognized" at 30 min, plus (b) words recalled at 30 min, but "not recognized" at 60 min	0–4	5–6	7–8	≥9
Exceedingly poor recognition Recognition of target words, summed across 30 and 60 min delays	>20	14–19	11–13	0–10

(Barrash, J., Surh, J., & Manzel, K., 2004)

Researchers hypothesized that participants simulating PTSD would perform significantly worse than compared to their genuine performances. Literature shows that people with PTSD do not perform significantly different than people from the general population. However, being unaware of this, it was expected that participants simulating the disorder would perform significantly worse compared to their baseline performances.

In summery, people with PTSD do not typically develop neuropsychological deficits. If such deficits do happen to occur, they are usually minor. Therefore, people who genuinely develop the disorder should not perform significantly different than people from the general population. However, those who feign the disorder often do not realize this. In their attempt to come off as having the disorder, they overplay the symptoms so drastically that their neuropsychological profile more closely resembles someone with traumatic brain injury or cognitive deficits. In such a case, when their performance is so poor, it is likely that they are feigning the disorder.

CHAPTER II

METHODS

2.1 Participants

Participants in this study were undergraduate and graduate college students from an urban Midwestern college who volunteered to participate in return for extra credit for psychology course work. In order to increase external validity, volunteers from the general population also participated. Participants ranged in age from 18-64 years. Exclusion criteria include any participants, or any participant’s family members who have been diagnosed with PTSD or any other axis I psychological disorder. Table III presents the descriptive characteristics of the current sample.

TABLE III: Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Age	58	46	18	64	27.33	13.080
Edu	58	7	12	19	14.31	1.993
Gender	N	%				
Male	29	50				
Female	29	50				

2.2 Procedure

The study is modeled after published dissimulation studies by Elhai, Gold, Sellers, and Dorfman (2001), and Liljequist, Kinder and Schinka (1998). Similar procedures were used for the present study. Participants were randomly assigned to one of three administration groups. Refer to Table IV for administration procedures. Each group was administered the Digit Span, Digit Symbol, Trail Making Test and Rey Auditory Verbal Learning Test. The first two groups were administered all of the tests and asked to perform normally.

Researchers made every effort possible in order to educate the participants about PTSD. It is believed that the more educated and familiarized the participants are about how to simulate PTSD, the more realistic the study would be. Participants viewed an informative video on the symptoms of PTSD and given a hypothetical situation asking them to imagine that they were in a car accident and must respond to items on the following tests as if they had PTSD and were seeking financial compensation. The scenario includes a cautionary statement concerning the tests' ability to detect responses that are not consistent with PTSD, a list of symptoms that are consistent with PTSD, as well as the DSM-IV-TR criteria for PTSD.

While simulating PTSD, participants were given all of the tests previously mentioned, as well as responding to two questionnaires: the Impact of Events, a scale involving the severity of PTSD symptoms, and the Traumatic Event Inventory (TEI), a new scale that is designed to detect malingering in individuals who claim to have Post-Traumatic Stress Disorder (PTSD). The first two groups were counterbalanced in order to account for test-retest reliability. Group two was administered the test battery while

TABLE IV: Test Battery Administration Procedures For Testing Groups		
Group 1 (Baseline/Simulation group)	Group 2 (Simulation/Baseline group)	Group 3 (No Simulation group)
RAVLT Digit Span Digit Symbol TMT RAVLT-30 min delay Video Instructed to Simulate PTSD RAVLT Questionnaires RAVLT-30 min delay Digit Span Digit Symbol TMT RAVLT-60 min delay	Video RAVLT Questionnaires RAVLT-30 min delay Digit Span Digit Symbol TMT RAVLT-60 min delay Instructed to perform normally RAVLT Digit Span Digit Symbol TMT RAVLT-30 min delay	RAVLT Digit Span Digit Symbol TMT RAVLT-30 min delay Instructed to perform their absolute best RAVLT Digit Span Digit Symbol TMT RAVLT-30 min delay

performing normally after they had already taken the test simulating PTSD. Therefore, they had an advantage over the other group while measuring their baseline performance. Similarly, group one had an advantage because they simulated PTSD after they had already been exposed to the test battery during their baseline test administration.

A third group was given the same tests as the others, but instead of being asked to simulate PTSD during one of the test administrations, they were asked to retake the test while attempting to perform even better than previously. The purpose of this was to establish a control group for the previous two groups in order to demonstrate that there was no effect for repeatedly receiving the same test battery. Data from group three allowed researchers to measure test-retest scenarios.

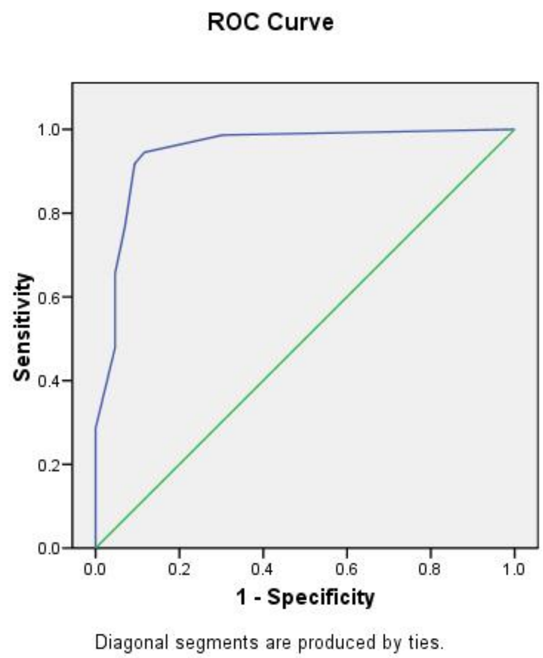
CHAPTER III

RESULTS

3.1 The Digit Span Test

Participants mean score on the Reliable Digit Span was significantly higher when they were performing genuinely (10.93) than when they were told to simulate PTSD (6.67). These scores were significantly different (p=.012). Utilizing a cut-off score of 7.5, this measurement was found to have good sensitivity (.94.5) and specificity (.884). Therefore, researchers decided to use this measurement as a 'gold standard,' in order to show that new indices have construct validity. Figure 1 displays the receiver operating characteristic (ROC) curve of the reliable digit span. The ROC curve displays the true positives (sensitivity) on the y-axis and false positive (1-specificity) on the x-axis.

Figure 1: ROC Curve of the Reliable Digit Span



3.2 The Digit Symbol Test

Mean number of symbols completed were significantly higher when participants were performing genuinely than when they were simulating PTSD. Participants performing genuinely completed on average of 79.25 symbols, whereas those simulating PTSD completed an average of 53.43. The Digit Symbol total score was found to have a high rate of sensitivity and specificity. Utilizing a cut-off value of 66.50, this measurement was found to have a sensitivity rate of 83.6% and specificity rate of 76.2%.

The Digit Symbol scaled score revealed a significant difference between groups ($p<.001$). The mean scaled score for participants simulating PTSD was 7.43, while the mean scaled score for participants performing genuinely was 14.14. This measure obtained an 84.7% sensitivity rate and an 81.0% specificity rate when implementing a

10.5 cut off score. In addition, the Digit Symbol scaled score significantly correlated with the RDS (r=.728, p<.01), verifying that this measure is an accurate indicator of effort.

Upon analysis, the Decomposition Paradigm (Poreh, 2000) was used in order to investigate implicit components and aid researchers in the development of new indices. Although all segments that were decomposed revealed significant differences between groups, the first 15 seconds produced the most interesting results. This segment revealed the largest mean difference in the number of symbols completed between the two groups (4.979). In addition, this measurement yielded good sensitivity (.861) and specificity (.810) when researchers used a cut-off rate of 11.50. This portion of the Digit Symbol test also significantly correlated with the RDS (r=.698, p<.01).

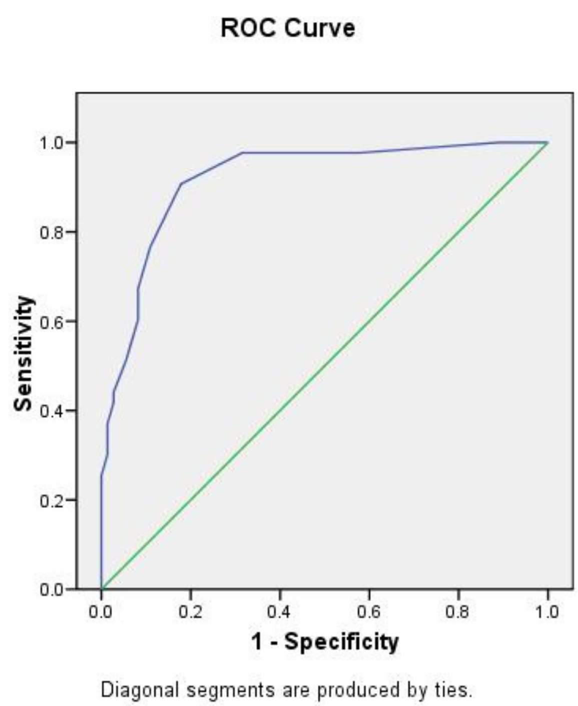
Table V: Digit Symbol Group Statistics

	Groups	N	Mean No. Completed	Std. Deviation
Time 1-15 (s)	Baseline	73	14.48	2.724
	Simulators	42	9.50	3.062
Time 16-30 (s)	Baseline	73	12.48	2.921
	Simulators	42	8.33	3.552
Time 31-45 (s)	Baseline	73	11.96	2.584
	Simulators	42	8.14	3.324
Time 46-60 (s)	Baseline	73	12.07	2.835
	Simulators	42	7.93	3.403
Time 61-75 (s)	Baseline	73	12.08	2.515
	Simulators	42	7.55	3.038
Time 76-90 (s)	Baseline	72	11.01	2.846
	Simulators	42	6.90	3.968
Total Time (s)	Baseline	73	79.25	11.166
	Simulators	42	53.43	20.853

3.3 The Trail Making Test

Participants simulating PTSD required significantly more time in order to complete the test than participants performing genuinely. Table VI displays the mean amount of time each group required in order to complete each section of the TMT. Both parts of the test have been decomposed into five second sections in order to investigate implicit components of the test that may be used to distinguish performance effort. Trails

Figure 2: ROC Curve for Trails 21-25 of the TMTA



in part A, section 21-25 were found to be particularly accurate in making this distinction. Researchers believe that this part of the test may be particularly accurate because it is the last section of this part of the TMT, and therefore it is much easier for the average participant to complete. At this point in the test, most of the dots have already been connected and, for that reason, the participant has fewer dots to choose from to connect.

To establish construct validity and determine the relationship between this section of the TMT and the RDS, Pearson correlational analyses were performed. A negative and significant correlation exists between these two measures ($r=-.492$, $p<.01$). In addition, scores of 5.50 or more were associated with high sensitivity (.907) and specificity (.822) for this section of the TMT.

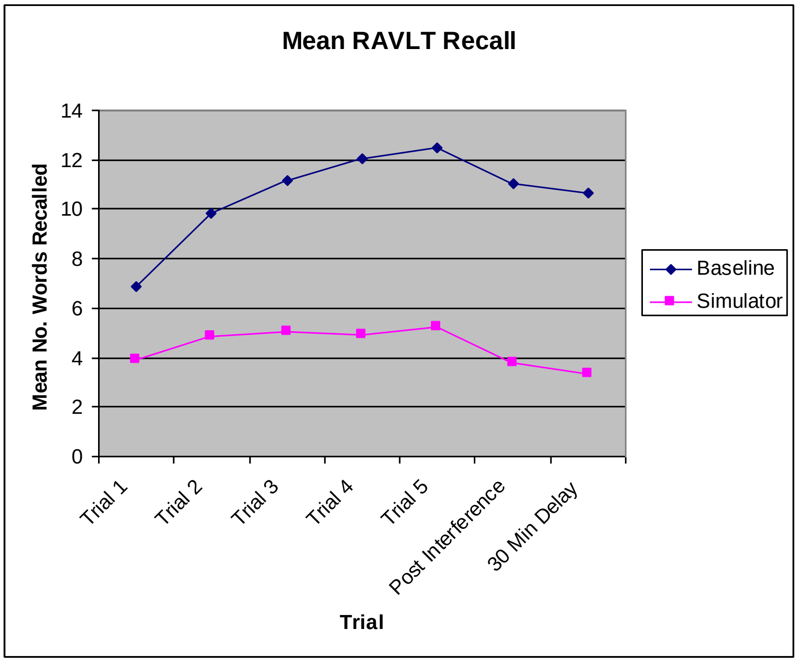
TABLE VI: Trail Making Test Group Statistics

	Group	N	Mean Time (secs)	Std. Deviation	Std. Error Mean
TMT Part A	Baseline	73	22.63	8.538	.999
	Simulators	43	69.00	82.942	12.648
TMT Part B	Baseline	73	53.38	20.054	2.347
	Simulators	43	117.44	93.974	14.331
TMT Part A Trails 1-5	Baseline	73	3.51	2.199	.257
	Simulators	43	11.70	14.620	2.230
TMT Part A Trails 6-10	Baseline	73	4.66	2.795	.327
	Simulators	43	12.42	17.073	2.604
TMT Part A Trails 11-15	Baseline	73	5.22	2.490	.291
	Simulators	43	16.86	21.906	3.341
TMT Part A Trails 16-20	Baseline	73	4.73	4.894	.573
	Simulators	43	14.74	23.459	3.578
TMT Part A Trails 21-25	Baseline	73	4.37	2.294	.269
	Simulators	43	13.23	11.785	1.797
TMT Part B Trails 1-5	Baseline	73	6.86	4.685	.548
	Simulators	43	18.00	16.355	2.494
TMT Part B Trails 6-10	Baseline	73	12.00	9.286	1.087
	Simulators	43	23.67	21.250	3.241
TMT Part B Trails 11-15	Baseline	73	12.89	6.865	.803
	Simulators	43	26.88	23.124	3.526
TMT Part B Trails 16-20	Baseline	73	11.37	5.606	.656
	Simulators	43	28.42	27.435	4.184
TMT Part B Trails 21-25	Baseline	73	10.26	4.773	.559
	Simulators	43	21.09	20.268	3.091

3.4 RAVLT

Figure 3 displays the results from the RAVLT. Participants simulating PTSD obtained a blunted curve on average when compared to those performing genuinely. In addition, simulators recalled an average of approximately seven fewer words during the post-interference trial; a statistically significant difference between the groups (p<.001). Lastly, the 30 Minute Delay portion of the test produced statistically significant results (p<.001). The approximate mean difference of words recalled on this portion of the test between the two groups was seven words.

FIGURE 3: RAVLT Learning Curve



The forced choice 60 minute delay portion of the test significantly correlated with the RDS (r=.430, p<.01). Findings on this measurement were limited because this portion of the RAVLT was only administered to participants while simulating PTSD. Researchers were not able to make as many conclusions with this measurement because there was no control group.

3.5 Construct Validity

The construct validity of newly developed neuropsychological tests were assessed by conducting correlational comparisons with established neuropsychological tests that have shown to validly distinguish malingerers from patients with genuine psychological disorders. These neuropsychological tests, such as the Reliable Digit Span, and the AVLTX, have served as a ‘gold standard’ in order to establish that the new tests presently under research are also valid indicators of malingering.

One index, derived from the TMT, that seemed to be most promising was the time it took participants to complete trails 21-25 on part A. Using the Pearson Correlation Coefficient, this index significantly correlated with the RDS ($p<.01$). Researchers believe that this particular section of the TMT was especially effective at measuring effort because these are the last five trails of Part A. By this point in the test, all other dots have been connected; so it is expected that most participants would be particularly faster at connecting the final five dots, because there are less choices to make. Table VII displays the correlations between trails 21-25 of the TMT part A and other validated indices of effort.

TABLE VII: Pearson Correlations between TMTA Trails 21-25 and Validated Measures of Effort

	Reliable Digit Span	Digit Symbol SS	Digit Symbol Est. Raw 120	RAVLT FC 30 Min
TMTA 21-25	-.492	-.554	-.590	-.563

3.6 Control group

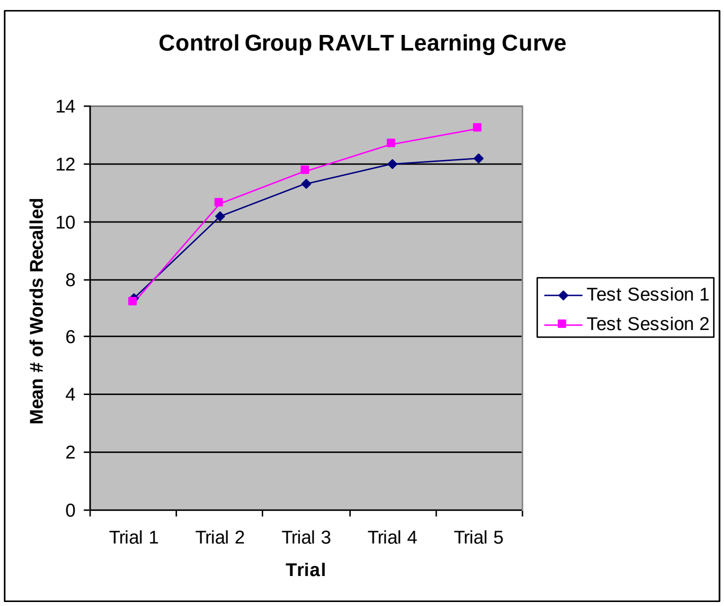
Scores from each test were analyzed to examine group effects. A comparison of means support the hypothesis that the control group did not perform significantly different between each test administration. These results indicate that participants are unable to perform significantly different despite practice effects and coaching. Therefore, there is no concern of a significant carryover effect from the first to the second test administration. Table VIII compares the mean scores on several indices that were obtained by the control group broken down by test administration session. There were no significant differences between test sessions.

Table VIII: Comparison of Means for Test Retest Reliability for Control Group

Group		N	Mean	SD	Sig.
Reliable Digit Span	Test Session 1	15	10.93	1.907	.429
	Test Session 2	15	11.47	1.727	
RAVLT Forced Choice 30 Min Delay	Test Session 1	15	14.60	1.056	1.00
	Test Session 2	15	14.60	.632	
Digit Symbol Total	Test Session 1	15	78.33	11.65	.165
	Test Session 2	15	83.73	8.836	
TMT B	Test Session 1	15	54.67	21.32	.547
	Test Session 2	15	50.20	18.67	

Additional analysis was made examining RAVLT learning curves broken down by test administration session. Figure 4 shows that the learning curves are not significantly different between test administration sessions. This contributes additional support that although there are minor differences, there is no sign cant carryover effect.

Figure 4: Control Group RAVLT Learning Curve



CHAPTER IV

DISCUSSION

The present study used a simulation design in order to develop new indices for distinguishing people feigning PTSD. Successfully developing these indices will give clinicians another tool, or indicator, in order to more accurately discern individuals who may have alternative motives for obtaining a diagnosis of PTSD. By creating additional indices, significant amount of money can be saved each year in unnecessary health care, social security, disability, insurance, and workers compensation costs. This, in turn would improve health care standards for those who genuinely have PTSD. In addition, resources would be better allocated in order to help those who genuinely have the disorder.

Overall, the results from this study support that indices developed from these four neuropsychological tests are accurate indicators of effort. These neuropsychological tests are great measures of effort, especially when feigning PTSD because this disorder does not significantly affect cognitive functioning. Many people do not realize this and assume that people with this disorder are unable to perform within average range. It is for this reason why people feigning the disorder drastically exaggerate, or even completely fabricate symptoms in order to exhibit impairment.

Each test successfully classified participants as either simulating PTSD or performing genuinely. The results obtained from this study demonstrate that these indices are accurately able to distinguish individuals and classify them into one of the two study groups. Of most interest were the results obtained from trails 21-25 of part A of the TMT. This index seems to be most promising in aiding clinicians to accurately identify people feigning PTSD. In addition to accurately classifying people who are or are not feigning PTSD, this index requires limited amount of time to administer. It is an easy measurement to administer. This index appears to be a great measurement for clinicians to utilize when they initially suspect poor effort.

It is important for clinicians to note that this index should not be used as an isolated diagnostic tool, but rather as an indicator for suspect effort. This index may serve as marker, or red flag, for people who may be feigning PTSD. Further testing is recommended in order to confirm clinician's hypothesis of poor effort.

In addition to this study, future research is suggested utilizing participants who genuinely have PTSD in order to further validate these findings. Future study participants should be obtained from litigating samples claiming to have PTSD, as well as non-litigating participants. It is understood by the researchers that the use of participants' simulating malingering may be a potential limitation of the study. Therefore, using participants obtained from a medico-legal context would increase validity.

It would also be necessary for future research to include data on different traumatic events from which PTSD can arise. For example, prevalence, as well as symptoms may differ for people claiming to have PTSD due to traumatic experiences such as rape, automobile accidents, or war. Therefore, it may be necessary to collect data utilizing samples from each group. In doing so, a more sophisticated profile could be developed that would specifically relate to individual subgroups.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and Statistical Manuel of Mental Disorders (5Ed).* Washington: American Psychiatric Press.

Axelrod, Bradley N., Fichtenberg, Norman L., Millis, Scott R., Wertheimer, Jeffrey C. (2004). Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale-third edition. *Clinical Neuropsychologist, 20*(3), 513-523.

Barrash, J., Surh, J., & Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (ALVTX): Validation with clinical samples. *Journal of Clinical and Experimental Neuropsychology, 26,* 125-140.

Binet, A., & Simon, T. (1905/1916). New methods for the diagnosis of the intellectual level of subnormals. In H.H. Goddard (Ed.), *Development of intelligence in children (the Binet-Simon Scale)* (E.S. Kite, Trans., pp. 37-90). Baltimore: Williams & Wilkins.

Boake, C. (2000). Historical note: Edouard Clapare`de and the Auditory Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 22,* 286–292.

Elhai, J.D., Gold, S.N., Sellers, A.H., Dorfman, W.I. (2001). The Detection of malingered posttraumatic stress disorder with MMPI-2 Fake Bad Indices. *Assessment, 8(2),* 221-236.

Frueh, B. C., Elhai, J. D., Gold, P. B., Monnier, J., Magruder, K. M., Keane, T. M., et al. (2003). Disability compensation seeking among veterans evaluated for posttraumatic stress disorder. *Psychiatric Services, 54*, 84 -91.

Goebel, R.A. (1983). Detection of faking on the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 39,* 731-742.

Guriel, J., Fremouw, W. (2003). Assessing malingered posttraumatic stress disorder: A critical review. *Clinical Psychology Review, 23*(7), 881-904

Galton, F. (1887). Supplementary notes on prehension in idiots. *Mind, 12*, 79-82.

Hall, Ryan C.W., Hall, Richard C.W. (2006). Malingering of PTSD: forensic and diagnostic considerations, characteristics of malingerers and clinical presentations. *General Hospital Psychiatry, 28*(6), 525-535.

Hebben, N. & Milberg, W. (2002). *Essentials of Neuropsychological Assessment.* New York: Wiley & Sons, Inc.

Heinly, Matthew T., Greve, Kevin W., Bianchini, Kevin J., Love, Jeffrey M., Brennan, Adrianne. (2005). WAIS digit span-based indicators of malingered neurocognitive dysfunction. *Assessment, 12*(4), 429-444.

Jacobs, J. (1887). Experiments on prehension. *Mind*. *12*, 75-79.

Johnson, Carrie. (2007, July 19). U.S. Targets Health-Care Fraud, Abuse: New Medicare Task Force Wins 3rd Conviction of 2007. Washington Post, p. D01.

Kaufman, Alan S., Lichtenberger, Elizabeth, O. (1999). *Essentials of WAIS-III Assessment*. New York: John Wiley & Sons, Inc.

Kaufman, Alan S., Lichtenberger, Elizabeth, O. (2002). *Assessing adolescent and adult intelligence* (2^{nd} ed.). Boston: Allyn and Bacon.

Kraiuhin, C., Gordon, E., Stanfield, P., Meares, R., & Howson, A. (1986). P300 and the effects of aging: Relevance to the diagnosis of dementia. *Experimental Aging Research, 12*(4), 187-192.

Lees-Haley, P.R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology, 53,* 745-755.

Lezak, M.D. (1995). *Neuropsychological Assessment (3rd Ed)*. Oxford University Press: New York.

Liljequist, L., Kinder, B.N., Schinka, J.A. (1998). An investigation of malingering posttraumatic stress disorder on the Personality Assessment Inventory. *Journal of Personality Assessment, 71*(3), 322-336.

Martin, T.A., Hoffman, N.M., & Donders, J. (2003). Clinical utility of the Trail Making Test ratio score. *Applied Neuropsychology*, *10*, 163–169.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Moritz, S., Heeren, D., Andersen, B., & Krausz, M. (2001). An analysis of the specificity and the syndromal correlates of verbal memory impairments in schizophrenia. *Psychiatry Research, 101*(1), 23-31.

Pankratz, L. & Binder, L.M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed.) (pp. 223-236). New York: The Guilford Press.

Petersen, R., Smith, G., Kokmen, E., Ivnik, R., & Tangalos, E. (1992). Memory functioning in normal aging. *Neurology, 42*, 396-401.

Poreh, Amir M. (2000). The Quantified Process Approach: An emerging methodology to neuropsychological assessment. *The Clinical Neuropsychologist, 14*, 212-222.

Reitan, R.M. & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Therapy and clinical interpretation.* Tucson: Neuropsychology Press

Resnick, P. J. (1997). Malingering of posttraumatic disorders. In R. Rogers (Ed.), Clinical assessment of malingering and deception (2nd ed.) (pp. 130–152). New York: Guilford Press.

Roger, R., Sewell, K.W., & Goldstein, A. (1994). Exploratory models of malingering: A prototypical analysis. *Law and Human Behavior*, *18,* 543-552.

Ryan, J.J., Brown, K.I., & Paolo, A.M. (2005). Discrimination of healthy and brain damaged elderly using the WAIS-R. *International Journal of Neuroscience, 23,* 140-145.

Ryan, Joseph J., Kreiner, David S. (2006). Clinical applications of the Digit Symbol-Coding subtest. In Amir M. Poreh (Eds.), *The Quantified Process Approach to Neuropsychological Assessment* (pp. 53-82). New York, NY: Taylor and Francis.

Ryan, J.J., Sattler, J.M., & Lopez, S.J. (2000). Age effects on Wechsler Adult Intelligence Scale-III subtests. *Archives of Clinical Neuropsychology*, *15*, 311-317.

Ryan, J.J., Schnakenberg-Ott, S.D., & Brown, K.I. (2004). *WAIS-III Digit Symbol-Coding patterns among healthy young adults.* Manuscript submitted for editorial review.

Sellers, A.H. & Nadler, J.D. (1992). A survey of current neuropsychological assessment procedures used for different age groups. *Psyhcotherapy in Private Practice*, *11,* 47-57.

Shum, D.H., Harris, D., & O'Gorman, J.G. (2000). Effects of severe traumatic brain injury on visual memory. *Journal of Clinical and Experiemntal Neuropsychology, 22*(1), 25-39.

Spreen, O., & Struass, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.) New York: Oxford University Press.

Starch, D. (1911). Experiments in educational psychology. New York: Macmillan.

Surh, J., Gunstad, J., Greub, B., & Barrash, J. (2004). Exaggeration Index for an expanded version of the Auditory Verbal Learning Test: Robustness to coaching. *Journal of Clinical and Experimental Neuropsychology, 26,* 416-427.

Taylor, S., Frueh, B.C., Asmundson, G. (2007). Detection and management of malingering in people presenting for treatment of posttraumatic stress disorder: Methods, obstacles, and recommendations. Journal of Anxiety Disorders, 21(1), 22-41.

Tombaugh, T. N. (1996). The test of memory malingering. Toronto, Canada: Multi-Health Systems.

U.S. War Department, Adutant General's Office. (1944). *Maual of the Army Individual Test.* Washington: Author.

Woodard, John L. (2006). Memory performance indexes for the Rey Auditory Verbal Learning Test. In Amir M. Poreh (Eds.), *The Quantified Process Approach to Neuropsychological Assessment* (pp. 53-82). New York, NY: Taylor and Francis.

Woodworth, R.S., & Wells, F.L. (1911). Association tests. Being a part of the Report of the Committee of the American Psychological Association on the Standardization of Procedure in Experimental Tests. *Psychological Monographs, 13* (Whole no. 57).

Zenderland, L. (1998). Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing. New York: Cambridge University Press.